

AD-A170 691

MEASURES OF DEPENDENCE FOR EVALUATING INFORMATION IN
CENSORED MODELS. (U) FLORIDA STATE UNIV TALLAHASSEE
DEPT OF STATISTICS M HOLLANDER ET AL. JUL 85

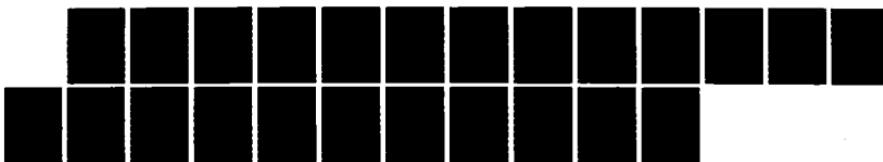
1/1

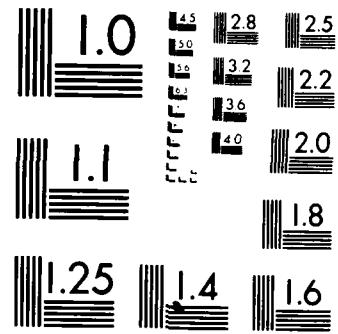
UNCLASSIFIED

FSU-STATISTICS-M786 AFOSR-TR-86-0547

F/G 12/1

ML





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

AD-A170 691

AFOSR-TR- 86-0547

(2)

Approved for public release,
distribution unlimited

Measures of Dependence for
Evaluating Information in Censored Models

by

Myles Hollander, Frank Proschan, and James Sconing

FSU Statistics Report M706

AFOSR Technical Report No. 85-180

July, 1985

DTIC

ELECTED

AUG 8 1986

B

The Florida State University
Department of Statistics
Tallahassee, Florida 32306-3033

Research sponsored by the Air Force Office of Scientific Research, AFSC, USAF, under Grant AFOSR-85-C-0007. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon.

DTIC FILE COPY

1981 AMS Subject Classification: Primary 62B10; secondary 62H20

Key words and Phrases: Censored model, coefficients of divergence, positive quadrant dependence, association.

080

86

BY CLASSIFICATION OF THIS PAGE

AD A170 691

REPORT DOCUMENTATION PAGE

REPORT NUMBER APOSR-TR. 86-0547		3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and subtitle) Measures of Dependence for Evaluating Information in Censored Models		5. TYPE OF REPORT & PERIOD COVERED Technical
6. AUTHOR(s) Myles Hollander, Frank Proschan, & James Sconing		7. PERFORMING ORG. REPORT NUMBER
8. CONTRACT OR GRANT NUMBER(s) F49620-85-C-0007		9. PRIMARY ELEMENT, PROJECT, TASK AREA, & WORK UNIT NUMBERS QMCJF 23C4 A5
10. CONTROLLING OFFICE NAME AND ADDRESS The U.S. Air Force Air Force Office of Scientific Research Bolling Air Force Base, DC 20332		11. REPORT DATE July, 1985
12. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		13. NUMBER OF PAGES 19
		14. SECURITY CLASS. (of this report)
		15a. DECLASSIFICATION/DEPROGRAMMING SCHEDULE

DISTRIBUTION STATEMENT (of this report)

distribution unlimited

DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from report.)

SUPPLEMENTARY NOTES

KEY WORDS

Censored model, coefficients of divergence, positive quadrant dependence, association.

ABSTRACT (Continue on reverse side if necessary and identify by block number)

Measures of information in censored models are developed by adapting measures of dependence between the lifetime variable and the observed variable. Some common notions of bivariate dependence and coefficients of divergence are used to derive these classes of measures. It is shown that most of the measures of bivariate dependence have the fundamental property that as censoring decreases stochastically, the information increases. An exception occurs when dependence is defined in terms of association. Conditions under which the coefficients of divergence enjoy the fundamental property are established.

Measures of Dependence for
Evaluating Information in Censored Models

by

Myles Hollander, Frank Proschan, and James Sconing

ABSTRACT

Measures of information in censored models are developed by adapting measures of dependence between the lifetime variable and the observed variable. Some common notions of bivariate dependence and coefficients of divergence are used to derive these classes of measures. It is shown that most of the measures of bivariate dependence have the fundamental property that as censoring decreases stochastically, the information increases. An exception occurs when dependence is defined in terms of association. Conditions under which the coefficients of divergence enjoy the fundamental property are established.

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
NOTICE OF TRANSMITTAL TO DTIC
This technical report has been reviewed and is
approved for public release IAW AFR 190-12.
Distribution is unlimited.
MATTHEW J. KERPER
Chief, Technical Information Division



A-1

1. Introduction.

Let X_1, X_2, \dots, X_n be independent and identically distributed (i.i.d.) random variables with distribution function (d.f.) $P(x) = \Pr(X \leq x)$. We consider the randomly right-censored model where the value of the random variable X is sometimes unobservable. Associated with each X_i is a variable Y_i independent of X_i . These Y_i 's are i.i.d. with d.f. $Q(x)$. The observations consist of the pairs (Z_i, δ_i) , $i = 1, 2, \dots, n$, where $Z_i = \min(X_i, Y_i)$, $\delta_i = I(X \leq Y_i)$, and $I(A)$ is the indicator function of the set A . Typically the concern of the statistician is how to best make use of the Z 's and δ 's to estimate F or some functional of F .

Given that this censoring is to take place another question arises. Suppose more than one censoring variable is available and the experimenter is given his choice as to which to use. Which variable should he choose? One approach is to choose the censoring variable which provides the greatest "information". Thus we seek general ways to measure information in censored models.

What properties should information measures possess? It is reasonable to expect that it is better to observe an X than a Y . Furthermore, stochastically increasing Y should increase information. Thus we consider the following two requirements for information measures.

If $Y_1 \stackrel{st}{\leq} Y_2$, the information in (Z_1, δ_1) is less than the information in (Z_2, δ_2) where $Z_i = \min(X, Y_i)$ and $\delta_i = I(X \leq Y_i)$, (1.1)
 $i = 1, 2$, for every X .

For every X and Y the information in X is greater than
 (1.2)
the information in (Z, δ) , $Z = \min(X, Y)$, $\delta = I(X \leq Y)$.

In general (1.2) will follow from (1.1) by taking $Y_2 \equiv \infty$ so that Z_2 has the same distribution as X . The adequacy of all information measures considered here will be with regard to (1.1) and (1.2). If (1.1) or (1.2) fail to hold, the measure is inadequate. Note that it is the monotonicity of the measure that is of interest as the measure can be made to increase or decrease by simply changing the sign in its definition.

In Section 2 notions of bivariate dependence are used to measure information. Models in which Y is increased stochastically should generally lead to increased dependence of X and Z . Thus measures of dependence provide a natural framework for studying information in the censored model.

Various notions of bivariate dependence are considered as candidates for measures of information. These include positive quadrant dependence (PQD), association, left-tail decreasing (LTD), right-tail increasing (RTI), and stochastically increasing (SI). Each of these is a notion of positive dependence which requires a certain probability that the variables are in some quadrant to be positive. These notions are extended to notions of "more positive quadrant dependent," "more associated," etc., by requiring that this probability be increasing. Then these new notions for increased positive dependence are considered for the role of measures of information. With properties (1.1) and (1.2) as criteria, it is shown that, with the exception of association, all of these notions of bivariate dependence are satisfactory.

In Section 3 the relationship between X and Z is explored through their related probability functions. Since Z is equal to X more often as censoring decreases it should be true that the probabilistic structure of Z should approach that of X as censoring decreases. One way to measure closeness of probability distributions is by coefficients of divergence. General classes of these measures

have been proposed independently by Csiszár (1963, 1967), Ali and Silvey (1965a, 1965b, 1966), and Ziv and Zakai (1973).

We use the following conventions: For a function f ,

$$\begin{aligned} f(0) &= \lim_{x \rightarrow 0} f(x) \\ 0 \cdot f\left(\frac{0}{0}\right) &= 0 \\ 0 \cdot f\left(\frac{a}{0}\right) &= \lim_{x \rightarrow \infty} a \frac{f(x)}{x}, \quad a > 0. \end{aligned} \tag{1.3}$$

Let $f(x)$ be a convex function. Let $\alpha(x)$ and $\beta(x)$ be nonnegative measurable functions on some measure space (X, \mathcal{A}, P) . Then the coefficient of divergence for $\alpha(x)$ and $\beta(x)$ is defined by

$$I_f(\alpha, \beta) = \int_X \beta(x) f\left\{\frac{\alpha(x)}{\beta(x)}\right\} dP(x). \tag{1.4}$$

For probability density functions $p_1(x)$ and $p_2(x)$, both absolutely continuous with respect to some measure λ , (1.4) becomes

$$I_f(p_1, p_2) = \int p_1(x) f\left\{\frac{p_2(x)}{p_1(x)}\right\} \lambda(dx) \tag{1.5}$$

This is the measure introduced by Csiszár. Ali and Silvey use a slightly different version defined by:

$$I_f(\phi) = E^*\{f(\phi)\} = \int_{\phi < \infty} f(\phi) dP_1 + P_2(N) \lim_{\phi \rightarrow \infty} \frac{f(\phi)}{\phi} \tag{1.6}$$

where ϕ is the generalized Radon-Nikodym derivative of P_2 with respect to P_1 and N is a P_1 -null set where P_2 has positive measure. Note that if p_1 and p_2 are mutually absolutely continuous, then (1.5) and (1.6) are identical.

For the censored model we need to find a satisfactory way to define p_1 and

p_2 in terms of X and Z . These must be designed with (1.1) and (1.2) in mind. In Theorem 3.2 we show that if p_1 and p_2 are taken to be the survival distributions of X and Z respectively, (1.1) and (1.2) are satisfied. It would seem more natural to let p_1 and p_2 be the respective densities of X and Z but Example 3.3 shows that this is unsatisfactory. However, if in (1.5) p_2 is taken to be the joint density of X and the vector $\underline{Z} = (Z, \delta)$, and p_1 is taken to be the product of the X and \underline{Z} marginals, then (1.1) holds with some restrictions on the convex function $f(x)$, and the density function of X , $p(x)$. Property (1.2) holds without any restrictions.

2. Measures of Bivariate Dependence.

Dependence measures have typically been developed to test for independence between two variables or to measure the degree to which large values of one variable go with large values of the other. Some general notions of dependence are given in the following definition.

Definition 2.1. Given two random variables U and V we say that U and V are:

- 1) Positively quadrant dependent (PQD) if $\Pr(U \leq u, V \leq v) \geq \Pr(U \leq u)\Pr(V \leq v)$ for all u, v . (2.1)

- 2) Associated if $\text{Cov}\{\Gamma(U, V), \Delta(U, V)\} \geq 0$, for all Γ, Δ which are componentwise increasing. (2.2)

- 3) Left-Tail Decreasing (LTD($V|U$)) if $\Pr(V \leq v|U \leq u)$ is decreasing in u . (2.3)

- 4) Right-tail Increasing (RTI($V|U$)) if $\Pr(V > v|U > u)$ is increasing in u . (2.4)

5) Stochastically Increasing ($SI(V|U)$) if $\Pr(V > v|U = u)$ is increasing in u . (2.5)

These notions are ordered in strength by:

$$SI(V|U) \Rightarrow RTI(V|U) \Rightarrow \text{Association} \Rightarrow PQD. \quad (2.6)$$

The sequence of implications is the same when $RTI(V|U)$ is replaced by $LTD(V|U)$. For verification of the implications and counterexamples to the reverse implications, see Barlow and Proschan (1975). Most of the above definitions were originally given in Lehmann (1966). The notion of association was introduced in Esary, Proschan, and Walkup (1967).

The inequalities in (2.1) - (2.5) are notions of positive dependence for a pair of variables. Next we compare the dependences of two sets of variables, specifically, between the variables X and Z_1 and X and Z_2 where $Z_i = \min(X, Y_i)$, $i = 1, 2$. For this a slight generalization of Definition 2.1 is needed.

Definition 2.2. Given four random variables U_1, U_2, V_1, V_2 , form two pairs of variables $\underline{W}_1 = (U_1, V_1)$ and $\underline{W}_2 = (U_2, V_2)$. We say that:

1) \underline{W}_1 is more PQD than \underline{W}_2 if for all u, v ,

$$\begin{aligned} \Pr(U_1 \leq u, V_1 \leq v) - \Pr(U_1 \leq u)\Pr(V_1 \leq v) &\geq \\ \Pr(U_2 \leq u, V_2 \leq v) - \Pr(U_2 \leq u)\Pr(V_2 \leq v). \end{aligned} \quad (2.7)$$

2) \underline{W}_1 is more associated than \underline{W}_2 if

$$\begin{aligned} \text{Cov}\{\Gamma(\underline{W}_1), \Delta(\underline{W}_1)\} - \text{Cov}\{\Gamma(\underline{W}_2), \Delta(\underline{W}_2)\} &\geq 0, \\ \text{for all componentwise increasing functions } \Gamma, \Delta. \end{aligned} \quad (2.8)$$

3) W_1 is more LTD than W_2 if

$$\begin{aligned} \Pr(V_1 \leq v | U \leq u') - \Pr(V_1 \leq v | U \leq u) &\geq \\ \Pr(V_2 \leq v | U \leq u') - \Pr(V_2 \leq v | U \leq u) & \end{aligned} \quad (2.9)$$

for all $v, u' < u$.

4) W_1 is more RTI than W_2 if

$$\begin{aligned} \Pr(V_1 > v | U > u) - \Pr(V_1 > v | U > u') &\geq \\ \Pr(V_2 > v | U > u) - \Pr(V_2 > v | U > u') & \end{aligned} \quad (2.10)$$

for all $v, u' < u$.

5) W_1 is more SI than W_2 if

$$\begin{aligned} \Pr(V_1 > v | U = u) - \Pr(V_1 > v | U = u') &\geq \\ \Pr(V_2 > v | U = u) - \Pr(V_2 > v | U = u') & \end{aligned} \quad (2.11)$$

for all $v, u' < u$.

With this definition, comparisons in the censored model can be made.

Theorem 2.3. In the censored model the amount of positive quadrant dependence increases as censoring decreases stochastically. That is, if $Y_1 \stackrel{\text{st}}{\leq} Y_2$ and $Z_i = \min(X, Y_i)$, $i = 1, 2$, then (X, Z_2) is more PQD than (X, Z_1) .

Proof: Consider $\Pr(X \leq x, Z_1 \leq z) - \Pr(X \leq x)\Pr(Z_1 \leq z)$. There are two cases.

1) If $x \leq z$, then

$$\begin{aligned} \Pr(X \leq x, Z_1 \leq z) - \Pr(X \leq x)\Pr(Z_1 \leq z) \\ = \Pr(X \leq x) - \Pr(X \leq x)\Pr(Z_1 \leq z) = P(x)\{1 - K_1(z)\} \\ = P(x)\bar{K}_1(z) = P(x)\bar{P}(z)\bar{Q}_1(z), \end{aligned}$$

where $\bar{K}_1(z) = \bar{P}(z)\bar{Q}_1(z)$, the survival function of Z_1 .

2) If $x > z$, then

$$\begin{aligned} \Pr(X \leq x, Z_i \leq z) &= \Pr(X \leq x)\Pr(Z_i \leq z) \\ &= \Pr\{X \leq x, \min(X, Y_i) \leq z\} - \Pr(X \leq x)\Pr(Z_i \leq z) \\ &= \Pr(X \leq z) + \Pr(z \leq X \leq x, Y_i \leq z) - \Pr(X \leq x)\Pr(Z_i \leq z) \\ &= P(z) + \{P(x) - P(z)\} Q_i(z) - P(x) \{1 - \bar{P}(z) \bar{Q}_i(z)\} \\ &= \bar{Q}_i(z) \{P(z) - P(x) + P(x) \bar{P}(z)\} = \bar{Q}_i(z)P(z) \bar{P}(x). \quad || \end{aligned}$$

With Theorem 2.3 it is easy to construct a class of measures for which (1.1) and (1.2) hold by taking averages of increasing functions of these positive quadrants. The following theorem is an easy consequence of Theorem 2.3.

Theorem 2.4. For any increasing function ψ , $\int \psi(\Pr(X \leq x, Z \leq z) - \Pr(X \leq x)\Pr(Z \leq z)) dx dz$ will increase as censoring decreases stochastically.

Corollary 2.5. Cov (X, Z) increases as censoring decreases stochastically.

Proof: Cov (X, Z) = $\iint \{\Pr(X \leq x, Z \leq z) - \Pr(X \leq x)\Pr(Z \leq z)\} dx dz$ and so the result is immediate from Theorem 2.4. ||

Covariance is, of course, a well known measure of positive dependence. Many other such measures can also be shown to increase as censoring decreases stochastically. To show this, we state the following theorem.

Theorem 2.6. Let $(U_i, V_i^{(1)})$, $i = 1, \dots, n$, be independent and identically distributed. Let $(U_i, V_i^{(2)})$, $i = 1, \dots, n$ be independent and identically distributed with $(U_i, V_i^{(1)})$ more PQD than $(U_i, V_i^{(2)})$, $i = 1, \dots, n$. Let r, s be concordant functions, that is, both r and s monotonic in the same direction in each argument. Then $\{r(U_1, \dots, U_n), s(V_1^{(1)}, \dots, V_n^{(1)})\}$ is more PQD than $\{r(U_1, \dots, U_n), s(V_1^{(2)}, \dots, V_n^{(2)})\}$.

The proof is by induction along the lines of Theorems 1 and 2 of Lehmann (1966).

Corollary 2.7. Kendall's τ , Spearman's ρ_s , and Blomqvist's q all increase as censoring decreases stochastically.

Proof: Kendall's $\tau = \text{Cov}(\text{sign}(X_2 - X_1), \text{sign}(Z_2 - Z_1))$ and hence is increasing by Theorem 2.6 and Corollary 2.5. Spearman's $\rho_s = 3\text{Cov}(\text{sign}(X_2 - X_1), \text{sign}(Z_3 - Z_1))$ and is increasing by Theorem 2.6 and Corollary 2.5. Blomqvist's $q = 2\{\Pr(X > m_x, Z > m_z) + \Pr(X \leq m_x, Z \leq m_z)\} - 1$ where m_x and m_z are the medians of X and Z respectively. This reduces to $2\{\Pr(X > m_x, Z > m_z) - \Pr(X > m_x)\Pr(Z > m_z) + \Pr(X \leq m_x, Z \leq m_z) - \Pr(X \leq m_x)\Pr(Z \leq m_z)\}$, which (from Theorem 2.3) increases as censoring decreases stochastically. ||

So the simple notion of positive quadrant dependence has yielded a large class of measures which can be used in the censored model. It is reassuring to note that these include some of the well-known measures of dependence. The next notion in the chain of (2.6) is association.

In Example 2.8 we show that even though there is less censoring, association may decrease. This is counter to the theme of (1.1) and so association is inappropriate as a measure of information in the censored model.

Example 2.8. Let $\Gamma(X, Z_i) = I(X > x_i, Z_i > z_i)$, $\Delta(X, Z_i) = I(X > x_2, Z > z_2)$, $i = 1, 2$, and let $x_1 < x_2 < z_1 < z_2$. Then $\text{Cov}\{\Gamma(X, Z_i), \Delta(X, Z_i)\} = \bar{P}(z_2)\bar{Q}_i(z_2) - \bar{P}(z_1)\bar{Q}_i(z_1)\bar{P}(z_2)\bar{Q}_i(z_2) = \bar{P}(z_2)\bar{Q}_i(z_2)\{1 - \bar{P}(z_1)\bar{Q}_i(z_1)\}$. Choose P, Q_1, Q_2 so that $\bar{P}(z_1) = 1/2, \bar{Q}_1(z_1) = 1, \bar{Q}_2(z_1) = 1/2, \bar{P}(z_2) = 1/4, \bar{Q}_1(z_2) = 5/12, \bar{Q}_2(z_2) = 1/3$. Note that $\bar{Q}_1(z_i) \geq \bar{Q}_2(z_i), i = 1, 2$. Then $\text{Cov}\{\Gamma(X, Z_1), \Delta(X, Z_1)\} = 5/96$, and $\text{Cov}\{\Gamma(X, Z_2), \Delta(X, Z_2)\} = 6/96$.

Thus a chain of implications similar to (2.6) using (2.7) - (2.11) is not

possible. This leaves the last three notions: LTD, RTI, and SI.

Theorem 2.9. If $Y_1 \stackrel{\text{st}}{\leq} Y_2$, then

- (i) (X, Z_2) is more RTI than (X, Z_1) ,
- (ii) (X, Z_2) is more LTD than (X, Z_1) , and
- (iii) (X, Z_2) is more SI than (X, Z_1) .

Proof: i) Let $x' < x$. Then

$$\begin{aligned} & \Pr(Z > z | X > x) - \Pr(Z > z | X > x') \\ &= \{\Pr(X > z, Y > z, X > x) / \Pr(X > x)\} - \{\Pr(X > z, Y > z, X > x') / \Pr(X > x')\}. \end{aligned} \quad (2.12)$$

There are three cases to consider.

- 1) Let $x > x' > z$. Then (2.12) reduces to $\Pr(Y > z) - \Pr(Y > z) = 0$.
- 2) Let $x > z > x'$. Then (2.12) reduces to $\Pr(Y > z) - \{\Pr(X > z, Y > z) / \Pr(X > x')\} = \bar{Q}(z)[1 - \{\bar{P}(z) / \bar{P}(x')\}]$. This decreases as $\bar{Q}(z)$ decreases.
- 3) Let $z > x > x'$. The (2.12) reduces to $\bar{P}(z)\bar{Q}(z)[\{1/\bar{P}(x)\} - \{1/\bar{P}(x')\}] = \bar{P}(z)\bar{Q}(z)\{\bar{P}(x) \cdot \bar{P}(x')\}^{-1}\{\bar{P}(x') - \bar{P}(x)\}$, which decreases as \bar{Q} decreases.

The proofs for LTD and SI follow in an analogous fashion. ||

Now as in the positive quadrant dependence case, classes of measures of information can be generated with Theorem 2.9.

Theorem 2.10. Let ψ be an increasing function. Then

- (1) $\int_z \int_{x < x'} \psi\{\Pr(Z \leq z | X \leq x') - \Pr(Z \leq z | X \leq x)\} dx dx' dz$ is increasing as censoring decreases stochastically,
- (2) $\int_z \int_{x < x'} \psi\{\Pr(Z > z | X > x) - \Pr(Z > z | X > x')\} dx dx' dz$ is increasing as censoring decreases stochastically,

(3) $\int_z \int_{x < z} \psi(\Pr(Z > z | X = x) - \Pr(Z > z | X = x')) dx dz$ is increasing as censoring decreases stochastically.

3. Coefficients of Divergence.

When $X < Y$ we have $Z = X$. Since the variables X, Z are often equal, in some sense their underlying probabilistic structures should be similar. From Kullback (1959), coefficients which increase as two distributions become less similar are called coefficients of divergence.

Csiszár (1963, 1966) generalized the Kullback-Leibler information number in the following fashion. Let $f(x)$ be a convex function on R^+ satisfying (1.3).

Let u_1 and u_2 be two probability distributions on some measurable space (X, A) . Let λ be a measure on (X, A) such that u_i is absolutely continuous with respect to λ , $i = 1, 2$. Let p_i be the Radon-Nikodym derivative of u_i with respect to λ . Define

$$I_f(u_1, u_2) = \int p_1(x) f\left[\frac{p_2(x)}{p_1(x)}\right] \lambda(dx). \quad (3.1)$$

$I_f(u_1, u_2)$ is the f-divergence of u_1 and u_2 .

From a completely different point of view, Ali and Silvey (1965a, 1965b, 1966) and independently Ziv and Zakai (1973) obtain an expression similar to (3.1). Both pairs of authors consider coefficients which measure the distance between two probability measures. Ali and Silvey postulate four properties which they believe the coefficient $d(P_1, P_2)$ should satisfy:

- 1) $d(P_1, P_2)$ should be defined for all measures P_1 and P_2 in the sample space.

- 2) $d(P_1, P_2) \geq d(P_1 t^{-1}, P_2 t^{-1})$ for all measurable transformations
 $y = t(x)$.
- 3) $d(P_1, P_1) \leq d(P_1, P_2)$ for all P_2 , and if P_1 is singular with respect to P_2 , $d(P_1, P_2) \geq d(P_1, P_3)$ for all P_3 .
- 4) Let $\{P_\theta; \theta \in (a, b)\}$ be a family of distributions with densities $p_\theta(x)$ having monotone likelihood ratio in x . Then if $\theta_1 < \theta_2 < \theta_3$,
 $d(P_{\theta_1}, P_{\theta_2}) \leq d(P_{\theta_1}, P_{\theta_3})$.

With these four postulates they define the coefficients of divergence as:

$$d_f(P_1, P_2) = E^*[f(\phi)] = \int_{\phi < \infty} f(\phi) dP_1 + P_2(N) \lim_{\phi \rightarrow \infty} f(\phi)/\phi, \quad (3.2)$$

where $f(x)$ is a convex function, $\phi = \frac{dP_2}{dP_1}$, and N is a P_1 -null set where P_2 has positive measure. The only difference between (3.1) and (3.2) is the dominating measure λ . The two measures will be identical if P_1 and P_2 are mutually absolutely continuous. Note that the measures (3.1) and (3.2) are not symmetric in P_1 and P_2 . However if $g(x) = xf(\frac{1}{x})$ then $I_f(p_1, p_2) = I_g(p_2, p_1)$. Further g is convex if and only if f is convex. Define a new function $f^*(x) = f(x) + g(x)$; then the measure $I_{f^*}(p_1, p_2)$ will be symmetric.

With the criteria (3.1) and (3.2), measures of information in the censored model can be generated by carefully choosing P_1 and P_2 in terms of X and Z . Note that P_1 and P_2 need not be probability measures. It is enough that both be integrable functions and the dominating measure be sigma-finite. Then the following can be used for information in the censored models.

Definition 3.1. Let X and Y have support on the positive real line. Then the information in the censored model is defined as in (3.1) where

$p_1(x) = \Pr(X > x)$, the survival function of X and $p_2(x) = \Pr(Z > x)$, the survival function of Z .

Theorem 3.2. With p_1 and p_2 defined as in Definition 3.1, $I_f(p_1, p_2)$ increases as censoring decreases stochastically.

Proof: Property (4) of Ali and Silvey (1966) will be used. We establish a partial ordering by saying $\alpha_1 < \alpha_2$ if $Y_{\alpha_1} \stackrel{st}{\geq} Y_{\alpha_2}$. The minimum for α corresponds to the uncensored case. Let $x_1 < x_2$, and note

$$\begin{vmatrix} \bar{P}(x_1) & \bar{P}(x_2) \\ \bar{P}(x_1)\bar{Q}(x_1) & \bar{P}(x_2)\bar{Q}(x_2) \end{vmatrix} = \bar{P}(x_1)\bar{P}(x_2) \{\bar{Q}(x_2) - \bar{Q}(x_1)\},$$

which is negative for all $x_2 < x_1$. Thus the monotone likelihood ratio property holds. ||

With this definition both (1.1) and (1.2) follow. At first it would seem that the more natural choice for p_1 and p_2 in (3.1) would be the density functions of X and Z . Since $X = Z$ when $X < Y$ it seems reasonable to postulate that the density of Z should approach that of X as censoring decreases. The following example shows that this need not be the case.

Example 3.3. Let X be defined on the points $x_1, x_2, x_3, x_4, x_5, x_6, x_7$ with $\Pr(X = x_i) = 1/12$ for $i = 1, 2, 3, 5, 6, 7$, and $\Pr(X = x_4) = 1/2$. Define three censoring variables Y_1, Y_2, Y_3 , also with support on $\{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$, satisfying:

$$\begin{aligned} \Pr(Y_1 = x_i) &= 1/12, \quad i = 1, 2, 4, 5, 6, 7 & \Pr(Y_1 = x_3) &= 1/2, \\ \Pr(Y_2 = x_i) &= 1/12, \quad i = 1, 2, 3, 5, 6, 7 & \Pr(Y_2 = x_4) &= 1/2, \\ \Pr(Y_3 = x_i) &= 1/12, \quad i = 1, 2, 3, 4, 6, 7 & \Pr(Y_3 = x_5) &= 1/2. \end{aligned}$$

Note that $Y_1 \leq^{\text{st}} Y_2 \leq^{\text{st}} Y_3$. Let Y_i be independent of X , $i = 1, 2, 3$, and let Z_i be the censored variable associated with Y_i . It should be true that $I_f(X, Z_1) \geq I_f(X, Z_2) \geq I_f(X, Z_3)$, so that the least censored variable is the least divergent from X and the most censored variable is the most divergent.

We need to compute $I_f(X, Z_i)$. Note that X and Z_i are mutually absolutely continuous, so that $\lambda(x)$ in (3.1) is the counting measure. Also, if $X = Y$, we adopt the convention that a death has been observed. Then with this convention, direct calculations show that the vectors of probabilities $\{\Pr(Z_i = x_1), \dots, \Pr(Z_i = x_7)\}$ are for $i = 1, 2, 3$ respectively, $(23/144, 21/144, 64/144, 27/144, 5/144, 3/144, 1/144)$, $(23/144, 21/144, 19/144, 1/2, 5/144, 3/144, 1/144)$, $(23/144, 21/144, 19/144, 57/144, 20/144, 3/144, 1/144)$. Then direct substitution into (3.1) yields

$$I_f(X, Z_1) = (1/12)f(23/12) + (1/12)f(21/12) + (1/12)f(16/3) + (1/2)f(3/8) \\ + (1/12)f(5/12) + (1/12)f(3/12) + (1/12)f(1/12)$$

$$I_f(X, Z_2) = (1/12)f(23/12) + (1/12)f(21/12) + (1/12)f(19/12) + (1/2)f(1) \\ + (1/12)f(5/12) + (1/12)f(3/12) + (1/12)f(1/12),$$

$$I_f(X, Z_3) = (1/12)f(23/12) + (1/12)f(21/12) + (1/12)f(19/12) + (1/2)f(57/72) \\ + (1/12)f(20/12) + (1/12)f(3/12) + (1/12)f(1/12).$$

Thus,

$$I_f(X, Z_1) - I_f(X, Z_2) = (1/12)f(16/3) + (1/2)f(3/8) - (1/12)f(19/12) - (1/2)f(1),$$

$$I_f(X, Z_2) - I_f(X, Z_3) = (1/2)f(1) + (1/12)f(5/12) - (1/2)f(57/72) - (1/12)f(20/12).$$

Take $f(x) = x^2$. Then

$$I_f(X, Z_1) - I_f(X, Z_2) = 1.73 \geq 0.$$

$$I_f(X, Z_2) - I_f(X, Z_3) = -.0304 \leq 0.$$

The above inequality reverses the expected order. Why does this happen?

Note that X has a large mode at the point x_4 . In the censoring variables, Y_i , the mode moves from x_3 to x_4 to x_5 . As the mode of Y_i moves toward the mode of X , Z_i resembles X more and more. When the mode of Y_i reaches that of X , Z_i will resemble X closely; Z_i will be unimodal at the same point as X . Now as the mode of Y_i continues moving to the right, Z_i no longer has such large probability at x_4 . If the mode of Y_i continues to the right, eventually Z_i becomes bimodal and thus Z_i appears to be different from X by a substantially greater amount than when the modes are equal.

Thus Example 3.3 shows that with the choice p_1 the density of X , p_2 the density of Z , criterion (1.1) fails to hold. It is true however that for this choice of p_1 , p_2 the measure does satisfy (1.2). This follows immediately from Lemma 1.1 of Csiszár (1967) or from Property 3 of Ali and Silvey (1966).

In order to develop a more satisfactory measure we consider the vector $\underline{Z} = (Z, \delta)$ and return to the concept of dependence.

Consider X and \underline{Z} as the two variables of interest; then the f -divergence of the Radon-Nikodym derivative of the joint distribution of X and \underline{Z} with respect to the product of their marginals is the information measure.

Note that the joint density of X and \underline{Z} puts positive probability on the line where $X = Z$, the 45° line passing through the origin. This line has zero two-dimensional Lebesgue measure. Thus p_1 and p_2 defined as the joint distribution of X and \underline{Z} and the product of the marginals are not mutually absolutely continuous. Hence the measures in (3.1) and (3.2) are no longer equivalent. Equation (3.2) is now useful only if $\lim_{x \rightarrow \infty} f(x)/x$ is finite. Equation (3.1) requires a measure $\lambda(x)$ which dominates both the joint density of X and \underline{Z} and the product of the marginals. Let $\lambda(x)$ be the sum of two-dimensional Lebesgue measure and a measure u , which is Lebesgue measure on the 45° line, $\{(x, y) : x = y, x > 0, y > 0\}$.

For the joint probability measure of (X, \underline{Z}) , we write $\Pr\{X = x, \underline{Z} = (z, 0)\} = p(x)q(z)$, for $x > z$, 0 otherwise, and $\Pr\{X = x, \underline{Z} = (z, 1)\} = p(x)\bar{Q}(x)$, for $x = z$, 0 otherwise. Then (3.1) becomes

$$I_f(p_X p_{\underline{Z}}, p_{X \times \underline{Z}}) = \int_0^\infty p(x)p(x)\bar{Q}(x)f\left\{\frac{p(x)\bar{Q}(x)}{p(x)p(x)\bar{Q}(x)}\right\} dx + \int_x \int_{z < x} p(x)q(z)\bar{P}(z)f\left\{\frac{p(x)q(z)}{p(x)q(z)\bar{P}(z)}\right\} dz dx,$$

which reduces to,

$$\int_0^\infty p(x)p(x)\bar{Q}(x)f\{1/p(x)\} dx + \int_0^\infty q(x)\bar{P}(x)\bar{P}(x)f\{1/\bar{P}(x)\} dx. \quad (3.3)$$

Take $g(x) = xf(1/x)$; then (3.3) becomes

$$I_g(p_X p_{\underline{Z}}, p_{X \times \underline{Z}}) = \int_0^\infty p(x)\bar{Q}(x)g\{p(x)\} dx + \int_0^\infty q(x)\bar{P}(x)g\{\bar{P}(x)\} dx. \quad (3.4)$$

The expression in (3.4) can be viewed as a loss function. In this case the amount $g\{p(x)\}$ is lost when $X = x$ is observed. If a censored observation is observed at time x , the loss is $g(\bar{P}(x))$. Now if censoring increases stochastically, losses $g(\bar{P}(x))$ occur more frequently while losses $g(p(x))$ occur less frequently. The original premise was that increased censoring leads to decreased dependence so that the joint distribution should be closer to the product of the marginals. Thus the f-divergence should decrease as censoring increases. Thus if $g(\bar{P}(x)) \leq g(p(x))$ the monotone criterion in (1.1) is satisfied.

Equation (3.4) can be rewritten as

$$I_g(p_X p_{\underline{Z}}, p_{X \times \underline{Z}}) = \int_0^\infty q(z)[\int_0^z p(x)g\{p(x)\} dx + \bar{P}(z)g\{\bar{P}(z)\}] dz. \quad (3.5)$$

This measure is equivalent to a measure of information in the discrete case developed in Hollander, Proschan, and Sconing (1985). Now if censoring increases stochastically (3.5) should decrease. This is equivalent to the term

$\psi(z) = \int_0^z p(x)g(p(x))dx + \bar{P}(z)g(\bar{P}(z))$ being increasing. Assume g is differentiable, then

$$\psi'(z) = p(z)g(p(z)) - p(z)\bar{P}(z)g'(\bar{P}(z)) - p(z)g(\bar{P}(z)),$$

which is positive if and only if for every z

$$g(p(z)) \geq \bar{P}(z)g'(\bar{P}(z)) + g(\bar{P}(z)). \quad (3.6)$$

Unfortunately inequality (3.6) is not always satisfied. For example, take $g(x) = -\log x$ and $\bar{P}(x) = \exp\{-\lambda x\}$; then the direction of the inequality depends on λ . However some conditions can be found for $g(x)$ and $p(x)$ so that (3.6) is satisfied. Two such conditions are:

C1: g decreasing on $[0, 1]$ and $p(z)\{\bar{P}(z)\}^{-1} \leq 2$

C2: g increasing on $[0, 1]$ and $p(z)\{\bar{P}(z)\}^{-1} \geq 2$

Theorem 3.4. If either C1 or C2 hold and $g'(x)$ is continuous on $[0, \infty]$, then $I_g(p_X p_{Z|X}, p_{X \times Z})$ is decreasing as censoring increases stochastically.

Proof: It is enough to show (3.6). Expand $g(p(z))$ in a Taylor series about $\bar{P}(z)$. Then

$$\begin{aligned} g(p(z)) &\geq g(\bar{P}(z)) + g'(\bar{P}(z))\{p(z) - \bar{P}(z)\} \\ &\geq g(\bar{P}(z)) + \bar{P}(z)g'(\bar{P}(z)) + g''(\bar{P}(z))\{p(z) - 2\bar{P}(z)\} \\ &\geq g(\bar{P}(z)) + \bar{P}(z)g'(\bar{P}(z)), \end{aligned}$$

if $g''(\bar{P}(z))\{p(z) - 2\bar{P}(z)\} \geq 0$, which holds if C1 or C2 hold. ||

In terms of the original function $f(x)$, $g(x)$ decreasing is equivalent to $f(x)/x$ increasing, $1 \leq x < \infty$. Most of the functions $f(x)$ which are commonly used in f -divergence satisfy the necessary condition.

Example 3.5.

1) $f(x) = x \log x$	$g(x) = -\log x$	Kullback-Leibler Information number
2) $f(x) = (1/2)(x^{\frac{1}{2}} - 1)^2$	$g(x) = (1/2)(x^{\frac{1}{2}} - 1)^2$	Hellinger metric
3) $f(x) = (1/2) x - 1 $	$g(x) = (1/2) x - 1 $	city-block distance
4) $f(x) = (x - 1)^2$	$g(x) = (x - 1)^2/x$	χ^2 -distance

It is easy to verify that in the above four cases, $g(x)$ is decreasing. Note that the third function does not satisfy the conditions of Theorem 3.4. However the ordering still holds under slightly more restrictive conditions.

Theorem 3.6. If g is decreasing on $(0, 1)$ and $p(z)(\bar{P}(z))^{-1} \leq 1$, then

$I_g(p_{X \times \underline{Z}}, p_{X \times \underline{Z}})$ is decreasing as censoring increases stochastically.

Proof: If g is decreasing and $p(z)/\bar{P}(z) \leq 1$, $g\{p(z)\} \geq g\{\bar{P}(z)\}$. Equation (3.6) follows since $g'(x) \leq 0$ on $(0, 1)$. ||

These last two theorems use the divergence measure as defined in (3.1). As was stated previously (3.2) is not satisfactory unless $\lim_{x \rightarrow \infty} f(x)/x < \infty$. Of the four functions cited in Example 3.5 only the second and third functions fit this criterion. In particular the third function, $f(x) = (1/2)|x - 1|$ is the one originally proposed by Ali and Silvey (1965a) for measuring dispersion between the joint distribution of two variables and the product of their marginals. In the censored model, the set N corresponds to the set where $X = Z$, or equivalently, when $X \leq Y$. Then (3.2) becomes

$$d_f(p_{X \times \underline{Z}}, p_{X \times \underline{Z}}) = \int_0^\infty q(x) \bar{P}^2(x) f\{1/\bar{P}(x)\} dx + c \int_0^\infty p(x) \bar{Q}(x) dx \quad (3.7)$$

where $c = \lim_{x \rightarrow \infty} f(x)/x$.

Theorem 3.7. If f is such that $\lim_{x \rightarrow \infty} f(x)/x = c < \infty$ and $f(x)/x$ is increasing for $1 < x < \infty$, then $d_f(p_{X \times Z}, p_X p_Z)$ increases as censoring decreases stochastically.

Proof: Consider (3.7) as an expected loss over the variable Z with loss $\bar{P}(x)f\{1/\bar{P}(x)\}$ when $Z=x$ and $Y < X$, and loss c when $X \leq Y$. So the loss function can be written as $\bar{P}(x)f\{1/\bar{P}(x)\}I(Y < X) + cI(X \leq Y)$. As Y increases stochastically, so does Z . Since $f(x)/x$ increases to c as x increases, the loss function is increasing. Hence the expected loss increases. ||

In Example 3.5 both the Hellinger metric and the city-block distance satisfy the conditions of Theorem 3.7. The conditions in Theorem 3.7 are less restrictive than those of Theorem 3.4 in the sense that there is no condition on the distribution of X . Of course the conditions in Theorem 3.7 are more restrictive in the sense that they allow far fewer functions f .

Acknowledgement: We are grateful to Professor Ian McKeague for helpful conversations about information.

REFERENCES

Ali, S. M. and Silvey, S. D. (1965a). Association between Random Variables and the Dispersion of a Radon-Nikodym Derivative. J. Roy. Statist. Soc. Ser. B 27 100-107.

Ali, S. M. and Silvey, S. D. (1965b). A further Result on the Relevance of the Dispersion of a Radon-Nikodym Derivative to the Problem of Measuring Association. J. Roy. Statist. Soc. Ser. B 27 108-110.

Ali, S. M. and Silvey, S. D. (1966). A General Class of Coefficients of Divergence of One Distribution from Another. J. Roy. Statist. Soc. Ser. B 28 131-142.

Barlow, R. and Proschan, F. (1975). Statistical Theory of Reliability and Life Testing: Probability Models. New York: Holt, Rinehart, and Winston, Inc.

Csiszár, I. (1963). Eine Informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. Magyar Tud. Akad. Mat. Kutató Int. Közl 8 85-108.

Csiszár, I. (1967). Information-Type Measures of Difference of Probability Distributions and Indirect Observations. Studia Scientiarum Mathematicarum Hungarica 2 299-318.

Esary, J. D., Proschan, F., and Walkup, D. W. (1967). Association of Random Variables with Applications. Ann. Math. Stat. 38 1466-1474.

Hollander, M., Proschan, F., and Sconing. J. (1985). Information in Censored Models. Florida State University Department of Statistics Report M701.

Kullback, S. (1959). Information Theory and Statistics. New York: J. Wiley & Sons, Inc.

Lehmann, E. L. (1966). Some Concepts of Dependence. Ann. Math. Stat. 37 1137-1153.

Ziv, J. and Zakai, M. (1973). On Functionals Satisfying a Data Processing Theorem. IEEE Transactions on Information Theory IT-19 275-283.

E

N

D

D

T

I

C

9

-

86